COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO
80523

department of electrical engineering

(12)

AD A024535

ADAPTIVE BASEBAND DIGITAL COMMUNICATION
AND APPLICABLE CONVERGENCE RESULTS

Final Report
by
David C. Farden

ARPA Order Number:	1872
Contract Number:	10292750
Program Code No.:	5G10
Naval Undersea Center Contract Number:	N66001-75-C-0224
Principal Investigators:	David C. Farden (303) 491-6018 Louis L. Scharf (303) 491-5065
Name of Contractor:	Colorado State University
Project Engineer:	Darrell E. Marsh (714) 225-6135
Effective Date of Contract:	30 May 1975
Short Title of Work:	Communication Receiver Studies
Contract Expiration Date:	1 March 1976
Amount of Contract:	$14,998.00

Sponsored by
Advanced Research Projects Agency
ARPA Order No. 1872

See Form 1473

The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

BEST
AVAILABLE COPY

ABSTRACT

Several baseband adaptive communication systems are presented and analyzed with regard to "self-synchronization" properties, intersymbol interference, and convergence properties. Recent convergence results, the proofs of which are contained in a companion report, are applied to provide extremely mild "covariance decay-rate conditions" for which the algorithms treated converge with probability 1. Of special interest are the convergence results treating correlated cyclostationary training data. Recent results on maximum-likelihood sequence estimation are extended to treat the detection of general "nonlinearly modulated" digital data over linear dispersive channels and nonwhite additive noise. Adaptive techniques for training the new detector structure are proposed for use when the channel and/or the noise covariance function are unknown.

ACKNOWLEDGMENT

This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Naval Undersea Center, San Diego, under Contract Number N66001-75-C-0224

Table of Contents

ADAPTIVE BASEBAND DIGITAL COMMUNICATION AND APPLICABLE CONVERGENCE RESULTS

I. INTRODUCTION

The large amount of literature treating adaptive digital communication over an unknown dispersive linear channel covers a large number of adaptive schemes. Most of the proposed adaptive schemes make use of either a form of stochastic approximation algorithm or the least mean-square (LMS) algorithm [12] to train the weight vector of a transversal filter. Both types of algorithms can be called "stochastic gradient-following algorithms." The receiver structures that have been proposed using such transversal filters have run the range from adaptive linear equalizers to nonlinear decision-feedback equalizers to combinations of both under a wide variety of "optimality" criteria. The excellent review paper by Lucky [11] reviews the major structures and philosophies used.

Convergence properties of algorithms used for adaptive signal processing do not appear to have received nearly as much attention as the receiver structures themselves. In Chapter II, recent convergence results that are generally applicable to most algorithms proposed for adaptive signal processing applications are presented. The proofs of these convergence results are presented in a companion report [1].

In Chapter III, two adaptive receiver structures, developed at the Naval Undersea Center, San Diego, are presented. The adaptive direct channel modeller is shown to converge to a cyclical shift of the sum of contiguous baud-length portions of the channel unit pulse response. Such a property is highly desirable in that it leaves the detector performance invariant to phase differences in transmitter

and receiver clocks. The adaptive inverse channel modeller is also shown to have a kind of self synchronization feature. Convergence properties of several possible algorithms for each receiver are established using the results of Chapter II. Of special interest are the convergence results for cyclostationary training data, showing that one is not required to average the data over one period before iterating the algorithm.

Chapter IV contains a treatment of maximum-likelihood sequence estimation. This problem has been treated previously by Forney [9] for pulse amplitude modulation (PAM), duration-limited channels, and additive white Gaussian noise, and by Ungerboeck [10] for modulation schemes which can be treated as complex PAM (e.g., PAM, phase modulation, quadrature modulation), duration-limited channels, and colored Gaussian noise. Ungerboeck [10] also proposes adaptive procedures for use when the channel impulse response and the noise covariance function are unknown. Both Forney and Ungerboeck make use of some form of the Viterbi algorithm. Magee and Proakis [14] propose an adaptive decision-feedback receiver structure to estimate the discrete-time channel response and incorporate this estimate with the results of Forney. Qureshi and Newhall [15] propose the use of an adaptive equalizer in cascade with a fixed Viterbi detector. The adaptive equalizer shortens the duration of resulting intersymbol interference, and hence, can result in a large savings in computational requirements for the Viterbi detector. In Chapter IV of the present work, a maximum-likelihood sequence estimator is developed for general digital modulation schemes,

i.e., for each alphabet symbol, a different waveform is transmitted.

In Chapter IV, such schemes are called "nonlinear modulation" which is, perhaps, a misleading term. In any case, the reader should interpret "nonlinear modulation" as used here to denote modulation schemes that cannot necessarily be treated as complex PAM modulation. Decision-directed adaptive procedures for implementing adaptive maximum-likelihood sequence estimation for these "nonlinear" schemes when the channel pulse response and the noise covariance function are unknown are also developed in Chapter IV.

II. SUMMARY OF RECENT CONVERGENCE RESULTS FOR STOCHASTIC APPROXIMATION ALGORITHMS

In a recent technical report [1] the author presents sufficient conditions for the almost sure (a.s.) convergence of a family of stochastic approximation algorithms. The family of algorithms treated includes those which can be suggestively called "stochastic gradient-following algorithms." Most algorithms commonly proposed for use in adaptive signal processing applications are included in the family of algorithms treated in [1]. Notable exceptions are algorithms which employ a "constant gain sequence." The only convergence results known to the author that treat such algorithms when the training data is correlated are those of Daniell [2] and Kim and Davisson [3]. The interested reader is referred to [4]-[5] for a more complete discussion of these points. The author strongly feels that the most generally applicable convergence results for algorithms suitable for adaptive signal processing applications when the training data is correlated are those in [1]. In this chapter, the results of [1] which are considered to be more practically useful are presented. The proofs are contained in [1] and will not be repeated here.

A. Notation and Basic Assumptions

Let $\{W_n\}$, $W_n \varepsilon R^p$ satisfy the recursion

$$W_{n+1} = W_n + \mu_n(P_n - F_n W_n) \tag{2.1}$$

for $n = 1,2,\ldots$, where R^p denotes the real p-dimensional Euclidean space, $\{\mu_n\}$ is a nonincreasing sequence of positive constants, $\{P_n\}$, $P_n \varepsilon R^p$, is a sequence of random variables, and $\{F_n\}$ is a sequence of real symmetric nonnegative definite pxp random matrices. The desired convergence is that of W_n to $w_o = R^{-1}P$, where

$$R = \lim_{n\to\infty} n^{-1} \sum_{\ell=1}^{n} E(F_\ell), \tag{2.2}$$

$$P = \lim_{n\to\infty} n^{-1} \sum_{\ell=1}^{n} E(P_\ell), \tag{2.3}$$

and $E(\cdot)$ denotes statistical expectation. The type of convergence of W_n to w_o under consideration is almost sure (a.s.) convergence. If $W_n \xrightarrow{a.s.} w_o$, then $\Pr(\lim_{n\to\infty} W_n = w_o) = 1$, where $\Pr(A)$ denotes the probability of event A. Note that R and P are simply the "time averages" of $E(F_\ell)$ and $E(P_\ell)$, respectively. It is assumed that R is positive definite and that $n^{-1} \sum_{\ell=a+1}^{n+a} E(F_\ell)$ converges uniformly to R as $n\to\infty$ for all positive integers a. This uniform convergence requirement is easily satisfied if $E(F_\ell)$ is either constant or periodic. It is further assumed that the sequence $\{\mu_n\}$ is such that $0<\mu_{n+1} \leq \mu_n$ for $n=1,2,\ldots$, and $0 < \lim_{n\to\infty} n\mu_n < \infty$. The single exception to this final requirement on $\{\mu_n\}$ is in Theorem 8.

The norm of a pxp matrix A, denoted by $||A||$, is defined here by $||A|| = \max_{|w|=1} |w'Aw|$, where $w \varepsilon R^p$. The ijth element of A is denoted by $(A)_{i,j}$. In case A is symmetric and nonnegative definite, $||A|| = \lambda_{max}(A)$,

where $\lambda_{max}(A)$ is the maximum eigenvalue of A. The norm of $w \epsilon R^p$, denoted by $|w|$, is taken to be $|w| = (w'w)^{\frac{1}{2}}$.

The convergence results presented in Section II-B involve "covariance decay rate" conditions on the processes $\{F_n\}$ and $\{P_n\}$. The required covariance functions and relationships between them are presented here for ready reference. Define

$$C_k = P_k - F_k w_o, \tag{2.4}$$

$$\rho_F(k,\ell) = E(F_k F_\ell) - E(F_k)E(F_\ell), \tag{2.5}$$

$$\rho_P(k,\ell) = E(P_k' P_\ell) - E(P_k')E(P_\ell), \tag{2.6}$$

and

$$\rho_{PF}(k,\ell) = E(P_k' F_\ell) - E(P_k')E(F_\ell). \tag{2.7}$$

Then

$$\rho_C(k,\ell) = E(C_k' C_\ell) - E(C_k')E(C_\ell) \tag{2.8}$$

can be expressed as

$$\rho_C(k,\ell) = \rho_P(k,\ell) + w_o' \rho_F(k,\ell) w_o - \rho_{PF}(k,\ell) w_o - \rho_{PF}(\ell,k) w_o. \tag{2.9}$$

Define

$$\rho_{PFP}(k,\ell,n) = E((P_k' - E(P_k'))F_n^2(P_\ell - E(P_\ell))), \tag{2.10}$$

$$\rho_{FF}(k,\ell,n) = E((F_k - E(F_k))F_n^2(F_\ell - E(F_\ell))), \tag{2.11}$$

and

$$\rho_{PFF}(k,\ell,n) = E((P_k' - E(P_k'))F_n^2(F_\ell - E(F_\ell))). \tag{2.12}$$

Then

$$\rho_{FC}(k,\ell,n) = E((C_k' - E(C_k'))F_n^2(C_\ell - E(C_\ell))) \tag{2.13}$$

can be expressed as

$$\rho_{FC}(k,\ell,n) = \rho_{PFP}(k,\ell,n) + w_o' \rho_{FF}(k,\ell)w_o - \rho_{PFF}(k,\ell,n)w_o$$
$$-\rho_{PFF}(\ell,k,n)w_o. \qquad (2.14)$$

B. Generally Applicable Convergence Results

The reader is cautioned against assuming that all of the conditions given below are necessary for the a.s. convergence of W_n to w_o. Proper combinations of sufficient conditons are presented in the theorems below. All of the assumptions made in Section II-A will be assumed to hold throughout the remainder of this chapter, with the single exception of Theorem 8.

CONDITION A1. Define

$$\gamma_F(i_1,i_2,i_3,i_4) = \max_{\substack{|w|=1 \\ w\varepsilon R^p}} E\left(\prod_{q=1}^{4} w'(F_{i_q} - E(F_{i_q}))w\right) \qquad (2.15)$$

Define $p_k(a) = a + [k^\alpha]$, $\nu_1(a) = 1$, $\nu_{k+1}(a) = \nu_k(a) + p_k(a)$, and $J_k(a) = \{\nu_k(a), \nu_k(a)+1, \ldots, \nu_{k+1}(a)-1\}$, for k=1,2,..., where a is a positive integer, [ ] denotes integer part, and $0 \leq \alpha < 1$. Condition A1 is that

$$\sum_{k=1}^{\infty} p_k^{-4}(a) \sum_{i,j,\ell,m\varepsilon J_k(a)} \gamma_F(i,j,\ell,m)<\infty \qquad (2.16)$$

for some α, $0 \leq \alpha < 1$ and for some positive integer a.

CONDITION A2. There exists a real-valued nonnegative function f(i,j) such that for some $\beta > \frac{1}{4}$, $|u|^\beta f(k,k+u)$ is uniformly bounded for all nonnegative integers k and u, and such that

$$|\gamma_F(i,j,\ell,m)| \leq f^2(i,j)f^2(\ell,m) + f(i,j)f(i,\ell)f(j,m)f(\ell,m). \qquad (2.17)$$

CONDITION A3. For some $\nu > 0$,

$$u^\nu \max\{||\rho_F(k,k+u)||, |\rho_C(k,k+u)|, |\rho_{FC}(k,k+u,n)|\}$$

is uniformly bounded for all nonnegative integers k, u, and n.

CONDITION A4. The quantity

$$g_n = \sum_{k=n}^{\infty} \mu_k E(C_k) \tag{2.18}$$

exists and for some $\beta > 1$,

$$\sum_{n=1}^{\infty} |g_n|^{\beta} E(||F_n||^{\beta}) < \infty. \tag{2.19}$$

THEOREM 1. Suppose that the structure and basic assumptions of Section II-A are satisfied. If either Condition A1 or A2 is satisfied, and if both Conditions A3 and A4 are satisfied, then $W_n \xrightarrow{a.s.} w_o$ as $n \to \infty$.

CONDITION A5. The sequence $\{||F_n||\}$ is a.s. bounded (in n).

CONDITION A6. The quantity g_n given by (2.18) exists.

CONDITION A7. For some $\nu > 0$,

$$u^{\nu} \max \{||\rho_F(k,k+u)||, |\rho_C(k,k+u)|\}$$

is uniformly bounded for all nonnegativ integers k and u.

THEOREM 2. Suppose that the structure and basic assumptions of Section II-A are satisfied. If either Condition A1 or A2 is satisfied, and if Conditions A5, A6, and A7 are all satisfied, then $W_n \xrightarrow{a.s.} w_o$ as $n \to \infty$.

CONDITION A8. For some $\nu > 0$, $u^{\nu}|\rho_P(k,k+u)|$ is uniformly bounded for all nonnegative integers k and u.

CONDITION A9. The sequence $\{F_k\}$ is deterministic and the quantity

$$g_n = \sum_{k=n}^{\infty} \mu_k (E(P_k) - F_k w_o) \tag{2.20}$$

exists and $F_n g_n \to 0$ as $n \to \infty$.

THEOREM 3. Suppose that the structure and basic assumptions of

Section II-A are satisfied If Conditions A8 and A9 are satisfied, and if $\mu_n F_n \to 0$ as $n \to \infty$, then $W_n \xrightarrow{a.s.} w_o$ as $n \to \infty$.

CONDITION A10. The sequence $\{P_k\}$ is deterministic and the quantity

$$g_n = \sum_{k=n}^{\infty} \mu_k (P_k - E(F_k) w_o) \tag{2.21}$$

exists and for some $\beta > 1$,

$$\sum_{n=1}^{\infty} |g_n|^{\beta} E(||F_n||^{\beta}) < \infty. \tag{2.22}$$

CONDITION A11. For some $\nu > 0$, $u^{\nu} \max\{||\rho_F(k,k+u)||, ||\rho_{FF}(k,k+u,n)||\}$ is uniformly bounded for all nonnegative integers k,u, and n.

THEOREM 4. Suppose that the structure and basic assumptions of Section II-A are satisfied. If either Condition A1 or A2 is satisfied, and if both Conditions A10 and A11 are satisfied, then $W_n \xrightarrow{a.s.} w_o$ as $n \to \infty$.

THEOREM 5. Suppose that $\{F_n\}$ and $\{P_n\}$ are deterministic and that the structure and basic assumptions of Section II-A are satisfied. If Condition A9 is satisfied, and if $\mu_n F_n \to 0$ as $n \to \infty$, then $W_n \to w_o$ as $n \to \infty$.

Definition. A sequence of random variables is said to be M-dependent if for all index sets I and J, with $\min_{n \varepsilon I, m \varepsilon J} |n-m| > M$, the two sets of random variables $\{y_n : n \varepsilon I\}$ and $\{y_m : m \varepsilon J\}$ are statistically independent.

CONDITION A12. The quantities $\gamma_F(k,k,k,k)$, $\rho_P(k,k)$, $\rho_{PF}(k,k)$, $\rho_{PFP}(k,k,k)$, $\rho_{PFF}(k,k,k)$ are all bounded and the sequences $\{F_n\}$ and $\{P_n\}$ are M-dependent.

CONDITION A13. The quantity g_n given by (2.18) exists and $\sum_{n=1}^{\infty} |g_n|^4 < \infty$.

THEOREM 6. Suppose that the structure and basic assumptions of Section II-A are satisfied. If Conditions A12 and A13 are satisfied, then $W_n \xrightarrow{a.s.} w_o$ as $n \to \infty$.

C. Special Families of F_n and P_n

Let $\{X_j\}_{-\infty}^{\infty}$ be a sequence of R^p-valued zero-mean random variables and let $\{s_j\}_{-\infty}^{\infty}$ be a sequence of real-valued zero-mean random variables. Define $R_{xx}(k,\ell) = E(X_k X_\ell')$, $P_s(k,\ell) = E(s_k X_\ell)$, and $\rho_s(k,\ell) = E(s_k s_\ell)$. Assume that $R_{xx}(k,k+u)$, $P_s(k,k+u)$ and $\rho_s(k,k+u)$ are all periodic in k with period N for all integers u.

It is helpful in what follows to consider a "typical" problem in adaptive signal processing. Suppose that it is desired to choose $w \epsilon R^p$ to minimize

$$\xi(w) = N^{-1} \sum_{k=1}^{N} E((s_k - w'X_k)^2).$$

$$= N^{-1} \sum_{k=1}^{N} (\rho_s(k,k) - 2w'P_s(k,k) + w'R_{xx}(k,k)w) \tag{2.23}$$

$$= \sigma_s^2 - 2w'P + w'Rw,$$

where

$$R = N^{-1} \sum_{k=1}^{N} R_{xx}(k,k), \tag{2.24}$$

and

$$P = N^{-1} \sum_{k=1}^{N} P_s(k,k). \tag{2.25}$$

It is well known that if R is positive definite, then the desired solution is $w_o = R^{-1}P$. Assume now that R and/or P are unknown, and that it is desired to use algorithm (2.1), with F_n and P_n functions of the observed time series $\{X_j\}$ and $\{s_j\}$. Obvious candidates for F_n and

P_n which satisfy the basic structure required in Section II-A are

$$F_n = K_n^{-1} \sum_{j=n-K_n+1}^{n} X_j X_j' \tag{2.26}$$

and

$$P_n = K_n^{-1} \sum_{j=n-K_n+1}^{n} s_j X_j, \tag{2.27}$$

where K_n is a positive integer; e.g. $K_n = 1$, K, N, or n.

LEMMA 1. Let F_n and P_n be given by (2.26) and (2.27) and $K_n = K$, a constant. Suppose the entire sequence $\{\mu_k\}$ satisfies either $\mu_k = a([\frac{k}{N}] + b)^{-1}$ or $\mu_k = a(k+b)^{-1}$, where $a > 0$, $b \geq 0$, and [ ] denotes integer part. Suppose further that $||R_{xx}(k,k)||$ is bounded (in k). Then Condition A4 is satisfied (with $\beta=2$). (Note that Conditions A6 and A13 are also satisfied).

Note that conditions A1 through A4 can be interpreted as quite mild "covariance decay-rate" conditions on the sequences $\{F_n\}$ and $\{P_n\}$. Lemma 1 provides several choices of sequences $\{\mu_k\}$ in order to satisfy Condition A4. In case either N=1 or K_n=N, then $E(C_k) \equiv 0$, and Condition A4 may be replaced by the condition that $E(||F_n||^{\beta})$ is bounded (in n) for some $\beta > 1$. In view of the widespread use of algorithms fitting the framework of (2.1) with F_n and P_n given by (2.26) and (2.27)

and $K_n=K$, it is worthwhile to establish sufficient conditions directly on the sequences $\{s_j\}$ and $\{X_j\}$. Such conditions are readily established when $\{s_j\}$ and $\{X_j\}$ are joint discrete-time Gaussian random processes.

CONDITION A14. The sequences $\{s_j\}$ and $\{X_j\}$ are jointly normally distributed, and F_n and P_n are given by (2.26) and (2.27) with $K_n=K$ (a constant).

CONDITION A15. For some $\alpha > \frac{1}{4}$,

$$u^{\alpha} \max_{1\le i,j\le p} |(R_{xx}(k,k+u))_{i,j}|$$

is uniformly bounded for all nonnegative integers k and u.

CONDITION A16. For some $\nu > 0$, $u^{\nu} \max_{1\le i\le p} \{|\rho_s(k,k+u)|, |(P_s(k,k+u))_i|\}$

is uniformly bounded for all nonnegative integers k and u.

THEOREM 7. Suppose that F_n and P_n are given by (2.26) and (2.27) with $K_n=K$, and that the structure and basic assumptions of Section II-A are satisfied. If the conditions stated in Lemma 1 are satisfied, and if Conditions A14, A15, and A16 are satisfied, then $W_n \xrightarrow{a.s.} w_o$.

D. Remarks and Related Results

From a practical viewpoint, Theorem 7 is seemingly of great significance. In many signal processing applications, the "Gaussian" assumption is often well-founded, in view of the central limit theorems. In this case, for the family of algorithms represented by (2.1) with F_n and P_n given by (2.26) and (2.27) and $K_n=K$ (a constant), essentially all one needs to verify is that any scalar covariance function with lag u that one can compute for elements of $\{X_j\}$ decays more rapidly than $u^{-1/4}$, and that any scalar covariance function with lag u that one can compute for elements of $\{s_j\}$ and $\{X_j\}$ decays at least as rapidly as $u^{-\nu}$, for some

$\nu > 0$. It is indeed difficult to imagine any stochastic process having a bounded spectral density which does not possess this desired property.

The observant reader will have noticed that the family of algorithms represented by (2.1) with F_n and P_n given by (2.26) and (2.27) and $K_n=n$ has not yet been treated. In this case, it is reasonable to expect that $F_n \xrightarrow{a.s.} R$ and $P_n \xrightarrow{a.s.} P$ as $n \to \infty$. It is this case to which the following theorem is addressed. The extremely powerful results of Serfling [6]-[7] can be applied to establish conditions for which $F_n \xrightarrow{a.s.} R$ and $P_n \xrightarrow{a.s.} P$. See also the proof of Corollary (4.5) in [1].

<u>THEOREM 8.</u> Suppose that there exist sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ of nonnegative real numbers (possibly random) satisfying $||F_n-R|| \overset{a.s.}{\leq} a_n$ and $|F_n w_o - P_n| \overset{a.s.}{\leq} b_n$. Further, suppose that there exists a positive integer n_o (possibly random) such that for all $n \geq n_o$, $0 < \mu_n(\lambda_{min}(R)-a_n) \leq 1$, where $\lambda_{min}(R)$ denotes the minimum eigenvalue of R. Then for all $n \geq n_o$,

$$|W_{n+1}-w_o| \leq |W_{n_o}-w_o| \prod_{k=n_o}^{n} (1-\mu_k d_k) + \max_{n_o \leq k \leq n} (b_k/d_k)\left(1 - \prod_{j=n_o}^{n}(1-\mu_j d_j)\right), \tag{2.28}$$

where $d_k = \lambda_{min}(R)-a_k$. Furthermore, if $\sum_{k=1}^{n} \mu_k d_k \xrightarrow{a.s.} \infty$ and $b_n d_n^{-1} \xrightarrow{a.s.} 0$ as $n \to \infty$, then $|W_n-w_o| \xrightarrow{a.s.} 0$ as $n \to \infty$.

Finally, although the previous results have treated only R^p-valued algorithms, a simple trick can be used to apply the above results to complex-valued algorithms. Consider

$$\Gamma_{n+1} = \Gamma_n + \mu_n(G_n - H_n\Gamma_n), \tag{2.29}$$

where H_n is Hermitian non-negative definite. Using superscripts r and i to denote real and imaginary parts, respectively, it is easily shown that

$$\begin{bmatrix} \Gamma^r_{n+1} \\ \\ \Gamma^i_{n+1} \end{bmatrix} = \begin{bmatrix} \Gamma^r_n \\ \\ \Gamma^i_n \end{bmatrix} + \mu_n \left\{ \begin{bmatrix} G^r_n \\ \\ G^i_n \end{bmatrix} - \begin{bmatrix} H^r_n & - & H^i_n \\ & & \\ H^i_n & & H^r_n \end{bmatrix} \begin{bmatrix} \Gamma^r_n \\ \\ \Gamma^i_n \end{bmatrix} \right\} . \tag{2.30}$$

Consequently, complex-valued algorithms such as (2.29) with H_n Hermitian can be put into the form of (2.1) with F_n real and symmetric by making use of (2.30). Furthermore, it is easily shown that the resulting F_n is positive definite if and only if H_n is positive definite.

III. CHANNEL ADAPTIVE STRUCTURES

In this chapter, two adaptive communication systems developed at the Naval Undersea Center, San Diego, California, will be presented and analyzed.

A. Basic Communication Scheme and Notation

Consider the communication scheme illustrated in Figure 3.1.

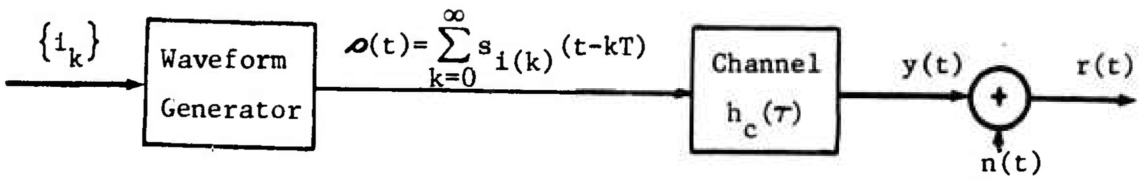

Figure 3.1. Generation of received process $\{r(t)\}$.

The information sequence $\{i_k\}$ is assumed to be composed of elements of the M-ary alphabet $\{1, 2,..., M\}$. During the kth baud interval, $t\epsilon[kT,(k+1)T)$, the waveform generator transmits the signal $s_{i(k)}(t-kT)$. It is assumed that $s_i(u) = 0$ for all $u\notin[0,T)$. Consequently, the transmitted signal, $s(t)$, can be expressed as

$$
\begin{aligned}
s(t) &= \sum_{k=0}^{\infty} s_{i(k)}(t-kT) \\
&= s_{i([\frac{t}{T}])}((t)_T),
\end{aligned}
\tag{3.1}
$$

where [ ] denotes largest integer part and $(t)_T$ = t modulo T. Note that $t = [\frac{t}{T}]\, T + (t)_T$. The output, y(t), of the linear time-invariant

channel, $h_c(\tau)$, is given by

$$y(t) = \int_{-\infty}^{\infty} s(t-\tau)h_c(\tau)d\tau. \tag{3.2}$$

The additive noise process, $\{n(t)\}$, is assumed to be zero-mean and wide-sense stationary, with autocorrelation $\rho_n(\tau) = E(n(t)n(t+\tau))$. It is also assumed that $\{y(t)\}$ and $\{n(t)\}$ are independent.

Assuming that $\{s(t)\}$ and $\{h_c(\tau)\}$ are "approximately" bandlimited to $f\varepsilon(-\frac{1}{2D}, \frac{1}{2D})$ and that $h_c(\tau)$ is "approximately" duration limited to $\tau\varepsilon[0,(L-1)D)$, i.e., that $h_c(\tau) \simeq 0$ for all $\tau\notin[0,(L-1)D)$, the continuous time model of Figure 3.1 can be approximately represented by the discrete-time system illustrated in Figure 3.2.

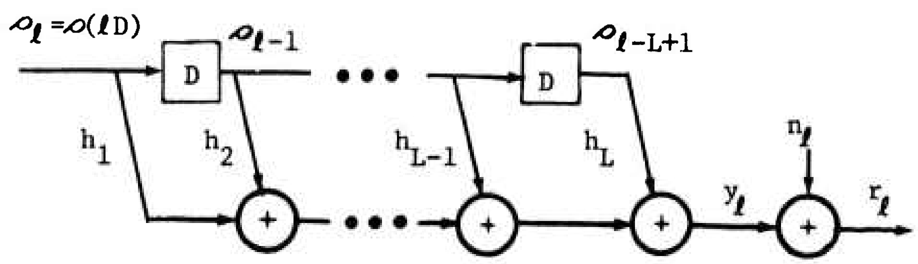

Figure 3.2. Discrete-time model of received process $\{r_\ell\}$.

In Figure 3.2, $h_j = h_c((j-1)D)$, $j=1,2,\ldots,L$. Consequently, the discrete-time received process, $\{r_\ell\}$, can be expressed as

$$r_\ell = y_\ell + n_\ell = \sum_{i=1}^{L} h_i s_{\ell-i+1} + n_\ell$$

$$= H' S_\ell + n_\ell \quad , \tag{3.3}$$

where

$$H = (h_1, h_2, \ldots, h_L)', \tag{3.4}$$

$$S_\ell = (s_\ell, s_{\ell-1}, \ldots, s_{\ell-L+1})', \tag{3.5}$$

and $'$ denotes matrix transpose. It is convenient to assume that $ND = T$, where N is a positive integer. Define the received data vector, R_ℓ, by

$$R_\ell = (r_\ell, r_{\ell-1}, \ldots, r_{\ell-N+1})'. \tag{3.6}$$

The conventional correlation receiver for estimating the information sequence $\{i_k\}$ is illustrated in Figure 3.3.

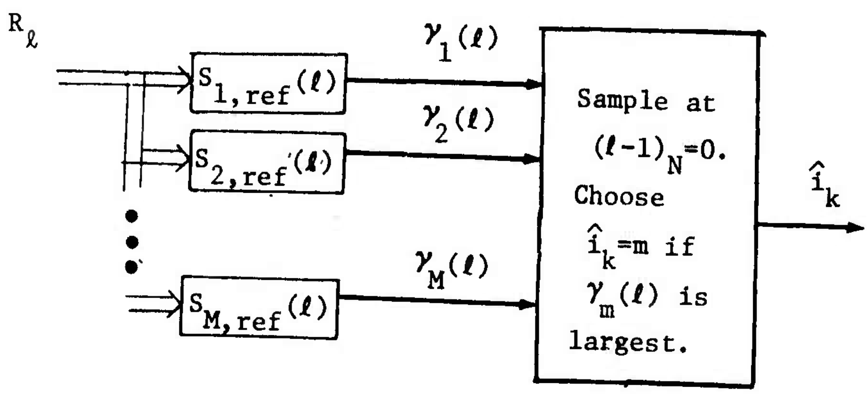

Figure 3.3. Conventional correlation receiver.

The set of statistics $\{\gamma_m(\ell) : 1 \leq m \leq M\}$ is determined from $\gamma_m(\ell) = R_\ell' S_{m,ref}(\ell)$, where R_ℓ is given by (3.6) and $S_{m,ref}(\ell)$ is an Nx1 reference vector, $m \in \{1,2,\ldots,M\}$. The set of reference vectors is a function of the channel, $h_c(\cdot)$, the basic signal set, $\{s_m(\cdot): m=1,2,\ldots,M\}$, and the noise autocorrelation function, $\rho_n(\cdot)$.

B. Direct Model

The direct model incorporates an adaptive receiver structure that establishes a transversal filter model for the communication channel. The basic idea is to apply a periodic reference signal, $s_{ref}((\ell)_N)$, composed of elements of the basic signal set as the input of an adaptive transversal filter having N taps and tap spacing D. The weight vector of the transversal filter is then trained to minimize the average mean-square error between the output of the transversal filter, z_ℓ, and the received data, r_ℓ. After some necessary notation is established, several highly desirable properties of this scheme are shown.

1. Notation and Basic Properties

Define the Nx1 vectors,

$$S_{ref}(\ell) = (s_{ref}((\ell)_N),\ s_{ref}((\ell-1)_N),\ \ldots,\ s_{ref}((\ell-N+1)_N))' \tag{3.7}$$

and

$$W = (w_1,\ w_2, \ldots,\ w_N)' \tag{3.8}$$

The output of the transversal filter, z_ℓ, can thus be expressed as

$$z_\ell = W'S_{ref}(\ell). \tag{3.9}$$

It is desired to find the W which minimizes

$$\xi(W) = \sum_\ell E((r_\ell - z_\ell)^2), \tag{3.10}$$

where the summation is over any N adjacent integer values of ℓ. Assuming that $\{i_k\}$ and $\{n_k\}$ are jointly wide-sense stationary, $E(y_\ell)$ is periodic (in ℓ) with period N. From (3.9), (3.10), and (3.3), we have

$$\xi(W) = \sum_\ell E(r_\ell^2) - 2W'S_{ref}(\ell)E(y_\ell) + W'S_{ref}(\ell)S'_{ref}(\ell)W. \tag{3.11}$$

The desired solution, w_o, is given as the solution to the set of linear equations

$$\sum_{\ell} S_{ref}(\ell)\, S'_{ref}(\ell) W = \sum_{\ell} S_{ref}(\ell) E(y_{\ell}). \tag{3.12}$$

Assuming that $\sum_{\ell} S_{ref}(\ell)\, S'_{ref}(\ell)$ is positive definite and that $\sum_{\ell} S_{ref}(\ell) E(y_{\ell}) \neq 0$, then a unique, nontrivial solution to (3.12) exists.

Assume that $\Pr\{i_k = m\} = p_m$ for $m = 1,2,\ldots,M$. Define

$S_m(\ell) = (s_m((\ell)_N D),\ s_m((\ell-1)_N D),\ldots,s_m(((\ell-N+1)_N D))'$

for $m = 1,2,\ldots, M$. Then

$$E(\delta_{\ell}) = \sum_{m=1}^{M} E(\delta_{\ell} | i_k = m) p_m \quad , \ kN \leq \ell < (k+1)N, \tag{3.13}$$

and hence

$$E(S_{i_k}(\ell)) = \sum_{m=1}^{M} S_m(\ell) P_m. \tag{3.14}$$

Let κ be a positive integer such that $(\kappa-1)N < L \leq \kappa N$ and define

$$H_i = (h_{iN+1}, h_{iN+2}, \ldots, h_{(i+1)N})', \ i = 0,1,\ldots,\kappa-1. \tag{3.15}$$

Recall that $h_{\ell} = 0$ for all $\ell > L$. From (3.3), (3.14), and (3.15) we have that

$$E(y_{\ell}) = H' E(S_{\ell}) = \sum_{i=0}^{\kappa-1} H_i' \sum_{m=1}^{M} p_m S_m(\ell). \tag{3.16}$$

Define a forward cyclical shift operator C_N such that for all $S = (s_1, s_2, \ldots, s_N)'$; $C_N S = (s_N, s_1, \ldots, s_{N-1})'$. It is easily seen that the NxN matrix

$$C_N = \begin{bmatrix} 0 & 1 & 0 & & & 0 \\ 0 & 0 & 1 & 0 & & 0 \\ & & & & & \\ & & & & & 0 \\ 0 & & & & & 1 \\ 1 & 0 & & & 0 & 0 \end{bmatrix} \tag{3.17}$$

satisfies the desired property. Suppose now that

$$S_{ref}(\ell) = \sum_{m=1}^{M} p_m \, S_m(\ell+n_d), \tag{3.18}$$

i.e., that the reference vector is $E(S_{i_k}(\ell+n_d))$.
Here n_d is a fixed but unknown shift between the transmitter and receiver clocks. From (3.17) and (3.18) we have

$$S_{ref}(\ell) = C_N^{n_d} \sum_{m=1}^{M} p_m S_m(\ell). \tag{3.19}$$

Consequently, (3.12) can be expressed as

$$C_N^{n_d} \sum_{\ell} \sum_{m=1}^{M} p_m S_m(\ell) \sum_{m_1=1}^{M} p_{m_1} S'_{m_1}(\ell) (C_N^{n_d})' W = C_N^{n_d} \sum_{\ell} \sum_{m=1}^{M} p_m S_m(\ell) \sum_{m_1=1}^{M} p_{m_1} S'_{m_1}(\ell) \sum_{i=0}^{\kappa-1} H_i . \tag{3.20}$$

Under certain obvious conditions on the $S_m(\ell)$, equation (3.20) provides the important result that

$$C_N^{n_d} W = \sum_{i=0}^{\kappa-1} H_i . \tag{3.21}$$

Result (3.21) shows that the solution to the minimization of (3.10) is a cyclical shift of $\sum_{i=0}^{\kappa-1} H_i$. If $\sum_{i=1}^{\kappa-1} H_i \approx H_0$ and an adaptive algorithm can be employed to converge to the solution of (3.21) without prior knowledge of n_d, then a self-synchronizing adaptive receiver strucutre will result. Note that the solution to (3.12) expressed by (3.21) suggests the presence of intersymbol interference in the resultant detector structure. It seems feasible to incorporate a baud-rate adaptive transversal filter to help alleviate the performance degradation due to this intersymbol interference.

2. Adaptive Direct Model

Consider a recursive algorithm of the form

$$W_{n+1} = W_n - \mu_n(F_n W_n - P_n), \tag{3.22}$$

for $n=1,2,\ldots$, and W_1 arbitrary. Algorithm (3.22) is identical with (2.1) and hence, convergence results presented in Chapter II are ideally suited to the adaptive direct model processor. In particular, note that algorithms of the form of (3.22) with

$$F_n = K_n^{-1} \sum_{j=n-K_n+1}^{n} S_{ref}(j)\, S'_{ref}(j), \tag{3.23}$$

$S_{ref}(j)$ given by (3.19) and

$$P_n = K_n^{-1} \sum_{j=n-K_n+1}^{n} r_j\, S_{ref}(j) \tag{3.24}$$

satisfy the specialized framework established by (2.1), (2.26), and (2.27). Of course, the assumption that $\{s_j\}$ and $\{X_j\}$ are zero mean, made in Section II-B, is violated. Furthermore, note that F_n given by (3.23) is deterministic, hence Theorem 3 is applicable.

The convergence results of Chapter II are now applied to treat the specialized family of algorithms represented by (3.22), (3.23), and (3.24). The assumptions and structure of the preceding subsection are assumed to hold throughout this subsection. In particular, it is assumed that w_o is the unique solution to (3.12) and that $S_{ref}(\ell)$ is given by (3.19). In other words, it is assumed that $w_o = R^{-1}P$, where

$$R = N^{-1} \sum_{j=1}^{N} S_{ref}(j)\, S'_{ref}(j), \tag{3.25}$$

$$P = N^{-1} \sum_{j=1}^{N} S_{ref}(j) E(y_j), \tag{3.26}$$

R is positive definite, and $P \neq 0$.

<u>CONDITION B1.</u> The entire sequence $\{\mu_k\}$ satisfies either

$\mu_k = a([\frac{k}{N}] + b)^{-1}$ or $\mu_k = a(k+b)^{-1}$, where $a > 0$, $b \geq 0$.

CONDITION B2. For some $\nu > 0$, $u^{\nu}|E(r_k r_{k+u}) - E(r_k)E(r_{k+u})|$ is uniformly bounded for all nonnegative integers k and u.

CONDITION B3. The sequence $\{\mu_k\}$ satisfies $0 < \mu_{k+1} < \mu_k$ and $0 < \lim_{k\to\infty} k\mu_k < \infty$.

THEOREM 9. If $K_n = K$ (a constant) in (3.23) and (3.24), and if Conditions B1 and B2 are satisfied, then $W_n \xrightarrow{a.s.} w_o$ as $n \to \infty$.

THEOREM 10. If $K_n = N$ in (3.23) and (3.24), and if Conditions B2 and B3 are satisfied, then $W_n \xrightarrow{a.s.} w_o$ as $n \to \infty$.

With these results established, it is worthwhile to note that Theorem 9 above establishes the desired convergence result for algorithms of the form of (3.22) that have the least demanding computational requirements, i.e., when $K_n = 1$. In this case, (3.22) can be expressed as

$$W_{n+1} = W_n - \mu_n S_{ref}(n)(z_n - r_n), \qquad (3.32)$$

where $z_n = S'_{ref}(n)W_n$ is the output of a transversal filter having the "weight vector" W_n and input $s_{ref}((n)_N)$. Even though Theorem 9 provides the desired convergence result for (3.32), other techniques for solving (3.12) should be considered, especially when convergence rate is an overriding issue.

3. Alternative Procedures for the Solution of (3.12)

Note that R given by (3.25) can be computed *a priori* and that

since $S_{ref}(j)$ is periodic with period N, R is a real symmetric Toeplitz matrix. Since efficient computational procedures are available for solving a Toeplitz system of equations (see e.g. [5]), one is lead to consider the following procedure:

$$\hat{P}_{n+1} = \hat{P}_n - \frac{1}{n+1}(\hat{P}_n - r_{n+1}S_{ref}(n+1)), \tag{3.33}$$

for n=1,2,..., and $P_1 = r_1 S_{ref}(1)$. Algorithm (3.33) is simply a recursive formulation of the usual sample mean estimator for P, i.e.,

$$\hat{P}_n = \frac{1}{n}\sum_{j=1}^{n} r_j S_{ref}(j). \tag{3.34}$$

Algorithm (3.33) also fits the framework of (2.1) by letting $F_n = I$, $W_n = \hat{P}_n$, and $P_n = r_{n+1}S_{ref}(n+1)$. In fact, $\hat{P}_n \xrightarrow{a.s.} P$ as $n \to \infty$ provided that Condition B2 above is satisfied. It is the author's conjecture that the rate of convergence of (3.33) is far more rapid than that of the previously considered algorithms. An estimate of w_o can be computed by solving

$$R\,W_n = \hat{P}_n$$

for W_n using techniques such as the Levinson algorithm or the Trench algorithm [5], [8].

A discussion of hardware implementations of such a procedure is given in [8]. Another technique is to use the Trench algorithm to compute R^{-1} and then estimate w_o by $W_n = R^{-1}\hat{P}_n$. A disadvantage to this approach is that R^{-1} may contain as many as $N^2/4$ distinct elements. Still another technique is to compute R *a priori* and use (3.22) with $F_n=R$ and $P_n = \hat{P}_n$ computed by (3.33). This final scheme makes efficient use of available *a priori* information and has only a moderate increase in computational requirements over (3.32). Note also that since R is an NxN symmetric Toeplitz matrix, only N elements of R need to be stored.

C. Inverse Model

The inverse model incorporates an adaptive transversal filter, the input of which is the discrete-time received sequence $\{r_\ell\}$. The "weight vector," W, of the transversal filter is trained to minimize the average mean-square error between the output of the transversal filter, z_ℓ, and a periodic reference signal, $s_{ref}((\ell)_N)$, composed of elements of the basic signal set. The name "inverse model" suggests that the adaptive transversal filter attempts to undo the distortion caused by the channel.

1. Notation and Basic Properties

Define the received "data vector" by

$$R = (r_\ell, r_{\ell-1}, \ldots, r_{\ell-N_1+1})', \tag{3.35}$$

and the transversal filter weight vector, W, by

$$W = (w_1, w_2, \ldots, w_{N_1})', \tag{3.36}$$

so that the output of the transversal filter at time ℓ is

$$z_\ell = W' R_\ell. \tag{3.37}$$

It is desired to choose W to minimize

$$\xi(W) = \sum_\ell E((z_\ell - s_{ref}((\ell)_N))^2), \tag{3.38}$$

where the summation is over any N adjacent integer values of ℓ. Assuming that $\{i_k\}$ and $\{n_k\}$ are jointly wide-sense stationary, $E(R_\ell)$, $E(R_\ell R_\ell')$, $E(z_\ell)$, and $E(z_\ell^2)$ are all periodic with period N.

Defining

$$R = N^{-1} \sum_{\ell=1}^{N} E(R_\ell R_\ell') \tag{3.39}$$

and

$$P = N^{-1} \sum_{\ell=1}^{N} \delta_{ref}\,((\ell)_N) E(R_\ell), \tag{3.40}$$

and assuming that R is positive definite and $P \neq 0$, the unique minimum of $\xi(W)$ is achieved when $W = w_o$, where

$$w_o = R^{-1}P. \tag{3.41}$$

It is assumed here that $\delta_{ref}((\ell)_N)$ is given by

$$\delta_{ref}((\ell)_N) = E(s_{i_k}((\ell)_N D))$$

$$= \sum_{m=1}^{M} E(\delta_\ell \,|\, i_k = m)p_m, \quad kN \leq \ell < (k+1)N. \tag{3.42}$$

The solution to the direct model, given by (3.21), is self-synchronizing and the relationship between the solution, w_o, and the channel unit pulse response is readily established. It is shown below that if $\delta_{ref}\,((\ell-n_c)_N)$ in place of $\delta_{ref}\,((\ell)_N)$ is used in (3.38) and (3.40), then the resulting solution, w_o^*, is not a simple cyclical shift of w_o given by (3.41). Results concerning the relationship between w_o and the channel unit pulse response are not readily obtained nor interpreted and hence, will not be included here.

Consider

$$\begin{aligned} P^* &= N^{-1} \sum_{\ell=1}^{N} \delta_{ref}\,((\ell-n_c)_N)E(R_\ell), \\ &= N^{-1} \sum_{\ell=1}^{N} \delta_{ref}((\ell)_N)E(R_{\ell+n_c}), \end{aligned} \tag{3.43}$$

where the last equality follows from the periodicity of $E(R_\ell)$ and a change of summation index. From (3.43), (3.17), and (3.40), it is

readily seen that

$$P^* = C_{N_1}^{n_c} P. \tag{3.44}$$

Consequently, the unique minimum of

$$\xi^*(W) = \sum_{\ell} E((z_\ell - s_{ref}((\ell - n_c)_N))^2) \tag{3.45}$$

is achieved when

$$W = w_o^* = R^{-1}P^* = R^{-1}C_{N_1}^{n_c} P = R^{-1}C_{N_1}^{n_c} R\, w_o. \tag{3.46}$$

Denote the ijth element of the real symmetric Toeplitz matrix R by $\gamma_{|i-j|}$, the ith element of P by $(P)_i$, the ith element of w_o by w_i and the ith element of w_o^* by w_i^*. Equation (3.46) implies that $R\, w_o^* = C_{N_1}^{n_c} P$ and $R\, w_o = P$. Consequently

$$\sum_{j=1}^{N_1} \gamma_{|i-j|}\, w_j^* = \sum_{j=1}^{N_1} \gamma_{|(i-n_c-1)_{N_1}+1-j|} w_j \tag{3.47}$$

for $i=1,2,\ldots,N_1$. For $n_c=1$ and $i=2$, equation (3.47) implies that

$$\gamma_1 w_1^* + \gamma_o w_2^* + \ldots + \gamma_{N_1-2} w_{N_1}^* = \gamma_o w_1 + \gamma_1 w_2 + \ldots + \gamma_{N_1-1} w_{N_1}. \tag{3.48}$$

Note that the left-hand side of (3.48) involves only $\gamma_o, \gamma_1, \ldots, \gamma_{N_1-2}$, while the right-hand side of (3.48) also involves γ_{N_1-1}. Consequently w_o^* cannot, in general, be a simple cyclical shift of w_o. However, a reasonable conjecture is that the resulting detector performance using w_o^* will not differ markedly from that using w_o provided that N_1 is large enough. Unfortunately, analytical justification for this conjecture is not available. Stochastic approximation algorithms for the solution of (3.41) are dealt with below.

2. Adaptive Inverse Model

Consider the algorithm

$$W_{n+1} = W_n - \mu_n (F_n W_n - P_n), \tag{3.49}$$

for $n=1,2,\ldots$, and W_1 arbitrary. Suitable choices for F_n and P_n are given by

$$F_n = K_n^{-1} \sum_{j=n-K_n+1}^{n} R_j R_j' \tag{3.50}$$

and

$$P_n = K_n^{-1} \sum_{j=n-K_n+1}^{n} \delta_{ref}((j)_N) R_j . \tag{3.51}$$

The family of algorithms represented by (3.49)-(3.51), of course, satisfies the general framework of Chapter II.

Convergence results of Chapter II are now applied to the family of algorithms represented by (3.41) with F_n and P_n given by (3.50) and (3.51). It is assumed that $w_o = R^{-1}P$, where R is given by (3.39) and P is given by (3.40). In case $\delta_{ref}((j-n_c)_N)$ is used in place of $\delta_{ref}((j)_N)$ in (3.51), all statements concerning the a.s. convergence of W_n to w_o must be replaced by equivalent statements concerning the a.s. convergence of W_n to w_o^*, where $w_o^* = R^{-1}P^* = R^{-1}C_{N_1}^{n_c}P$. Furthermore, it is assumed that $K_n = K$ (a constant) in (3.50) and (3.51). The case when $K_n = n$ is more readily handled by Theorem 8.

CONDITION C1. Define $f_i = r_i^2 - E(r_i^2)$, and $\gamma_r(\ell_1,\ell_2,\ell_3,\ell_4) =$

$$\max_{\substack{\ell_m - N_1 + 1 \le i_m \le \ell_m \\ m\epsilon\{1,2,3,4\}}} \left| E\left(\prod_{q=1}^{4} f_{i_q}\right)\right| .$$ Define $p_k = [k^\alpha]$,

$\nu_1=1$, $\nu_{k+1}=\nu_k+p_k$, and $J_k=\{\nu_k,\nu_k+1,\ldots,\nu_{k+1}-1\}$, for $k=1,2,\ldots$, and $0 \le \alpha < 1$. Condition C1 is that

$$\sum_{k=1}^{\infty} p_k^{-4} \sum_{i,j,\ell,m\varepsilon J_k} \gamma_r(i,j,\ell,m) < \infty \tag{3.52}$$

for some α, $0 \le \alpha < 1$.

Note that Condition A1 is satisfied if Condition C1 is satisfied and $K_n=K$ (a constant). Since $r_i = y_i + n_i$, it is of interest to establish conditions on the random sequences $\{y_i\}$ and $\{n_i\}$ which ensure the a.s. convergence of W_n to w_o. Recall that $\{y_i\}$ and $\{n_i\}$ are assumed to be independent and that $E(n_i) = 0$.

<u>LEMMA 2.</u> Define $g_i = y_i^2 - E(y_i^2)$ and $h_i = n_i^2 - E(n_i^2)$. If $E(n_{i_1} n_{i_2} \ldots n_{i_{2\ell-1}}) = 0$ for $\ell=1,2,\ldots,4$, then Condition C1 will be satisfied if

$$\sum_{k=1}^{\infty} p_k^{-4} \sum_{i,j,\ell,m\varepsilon J_k} \gamma_{gh}(i,j,\ell,m) < \infty, \tag{3.53}$$

where

$$\begin{aligned}\gamma_{gh}(i,j,\ell,m) = {} & |E(g_i g_j g_\ell g_m)| + |E(h_i h_j h_\ell h_m)| \\ & + |E(g_i g_j)||E(h_\ell h_m)| + |E(g_i g_j y_\ell y_m)||E(n_\ell n_m)| \\ & + |E(g_i y_\ell y_m)||E(h_j n_\ell n_m)| + |E(h_i h_j n_\ell n_m)||E(y_\ell y_m)| \\ & + |E(y_k y_j y_\ell y_m)||E(n_i n_j n_\ell n_m) ,\end{aligned} \tag{3.54}$$

and p_k, J_k are defined as in Condition C1.

Proof. The proof follows easily by expanding $E(f_i f_j f_\ell f_m)$ and collecting similar terms by subscript interchanges.

<u>LEMMA 3.</u> Let $\{g_i\}$, $\{h_i\}$, $\{n_i\}$, p_k, ν_k, and J_k be as in Lemma 2. Suppose $\{y_i\}$ is M-dependent, then Condition C1 will be satisfied if

$$\sum_{k=1}^{\infty} p_k^{-4} \gamma_{\nu_k, \nu_{k+1}-1} < \infty, \tag{3.55}$$

where

$$\gamma_{a,b} = \sum_{i=a}^{b} \sum_{j=a}^{b} \sum_{\ell=a}^{b} \sum_{m=a}^{b} \left(|E(h_i h_j h_\ell h_m)| + |E(h_i h_j n_\ell n_m)| + |E(n_i n_j n_\ell n_m)|\right)$$

$$+ \sum_{i=a}^{b} \sum_{\ell=a}^{b} \sum_{j=a}^{b} |E(h_j n_\ell n_i)| + (b-a) \sum_{i=a}^{b} \sum_{j=a}^{b} \left(|E(h_i h_j)| + |E(n_i n_j)|\right)$$

$$+ \sum_{i=a}^{b} \sum_{j=a}^{b} |E(h_j n_i^2)|. \tag{3.56}$$

Proof. It suffices to consider $\{y_i\}$ to be an independent, identically distributed sequence. Applying this fact to (3.54), the desired result follows.

Making use of Lemma 3 and techniques given in [1], Lemma 4 below can be proven.

<u>LEMMA 4.</u> If $\{y_i\}$ is M-dependent, if $\{n_k\}$ is jointly normally distributed, and if $u^{\beta}|E(n_k n_{k+u})|$ is uniformly bounded for all nonnegative integers k and u and some $\beta > 1/2$, then Condition C1 is satisfied.

Next consider Condition A3. With g_i and h_i as in Lemma 1,

$$r_i^2 - E(r_i^2) = g_i + h_i + 2y_i n_i, \tag{3.57}$$

and hence

$$E((r_i^2-E(r_i^2))(r_j^2-E(r_j^2)))=E(g_i g_j)+E(h_i h_j)+4E(y_i y_j)E(n_i n_j). \tag{3.58}$$

Making use of (3.58) and similar arguments for $\rho_P, \rho_{PF}, \rho_{PFP}, \rho_{FF}$, and ρ_{PFF}, the conditions stated in Lemma 4 can be shown to be sufficient for Condition A3. Theorems 11 and 12 can thus be readily established from these comments and Theorem 1.

Theorem 11. If the conditions of Lemma 4 are satisfied, if Condition B1 is satisfied, and if $K_n=K$ (a constant), then $W_n \xrightarrow{a.s.} w_o$ as $n \to \infty$.

Theorem 12. If the conditions of Lemma 4 are satisfied, if Condition B3 is satisfied, and if $K_n = N$, then $W_n \xrightarrow{a.s.} w_o$ as $n \to \infty$.

The condition that $\{y_k\}$ is M-dependent may or may not be a valid assumption in practice. If the information sequence $\{i_k\}$ is M_1-dependent, and the channel unit pulse response is duration limited, then $\{y_k\}$ is M-dependent. Theorem 1, of course, can also be applied to cases when $\{y_k\}$ is not M-dependent and $\{n_k\}$ is not a Gaussian random process.

IV. MAXIMUM-LIKELIHOOD SEQUENCE ESTIMATION FOR COMMUNICATION SYSTEMS USING NONLINEAR MODULATION SCHEMES

In recent years a large amount of literature has appeared treating nonlinear receivers for the detection of digital data transmitted over a dispersive linear channel. Theoretically, the most exciting development in this area is the application of the Viterbi algorithm to the solution of the problem of maximum-likelihood sequence estimation [9], [10]. The interested reader is referred to the excellent review paper by Lucky [11] for an overview of developments in this area. In this chapter, results are obtained for the maximum-likelihood sequence estimation of $\{i_k\}$ for the transmission model of Section III-A. The results obtained here are extensions of the work of Ungerboeck [10] to nonlinear modulation schemes.

A. Maximum-Likelihood Sequence Estimation

Recall that for the continuous time model of Section III-A, the received process $\{r(t)\}$ is given by $r(t) = y(t) + n(t)$. Assume that $\{n(t):-\infty<t<\infty\}$ is a zero-mean stationary Gaussian random process with $E(n(t)n(s)) = \rho_n(t-s)$. Assume that on the basis of the received data $\{r(t):\ 0\leq t\leq \ell T\}$ it is desired to find a maximum-likelihood estimate of $\{i_k\}_{k=1}^{\ell}$. Assume further that the channel, $h_c(\tau)$, as well as the noise covariance, $\rho_n(\tau)$, are known. Furthermore, assume that an inverse operator, ρ_n^{-1}, exists such that

$$\int_0^{\ell T} \rho_n^{-1}(t,s)\rho_n(t-s)ds = \delta(t), \tag{4.1}$$

where $\delta(t)$ is the dirac delta function. Note that, in general, ρ_n^{-1} depends on ℓT, and that ρ_n^{-1} cannot be interpreted as a linear

time-invariant system, even when $\{n(t)\}$ is stationary. The log-likelihood function, $\ell(r,\ell T)$, for $\{r(t): 0 \leq t \leq \ell T\}$ given $\{y(t):0\leq t\leq \ell T\}$ can be expressed as (within an additive constant and a positive multiplicative constant)

$$\ell(r,\ell T) = -\frac{1}{2}\int_0^{\ell T}\int_0^{\ell T} r(t)\rho_n^{-1}(t,s)r(s)dtds + \int_0^{\ell T}\int_0^{\ell T} r(t)\rho_n^{-1}(t,s)y(s)dtds - \frac{1}{2}\int_0^{\ell T}\int_0^{\ell T} y(t)\rho_n^{-1}(t,s)y(s)dtds. \quad (4.2)$$

The sequence $\{\hat{i}_k\}_{k=1}^{\ell}$ is considered to be the maximum-likelihood estimate of $\{i_k\}_{k=1}^{\ell}$ if $\{y(t):0\leq t\leq \ell T\}$ is the signal component of the received process corresponding to the information sequence $\{\hat{i}_k\}_{k=1}^{\ell}$ and this $\{\hat{i}_k\}_{k=1}^{\ell}$ results in the maximum of $\ell(r,\ell T)$ of any allowable sequence $\{i_k\}_{k=1}^{\ell}$. Since the maximization of $\ell(r,\ell T)$ is unaffected by the first term on the right hand side of (4.2), it is sufficient to maximize

$$\ell^*(r,\ell T) = \int_0^{\ell T}\int_0^{\ell T} r(t)\rho_n^{-1}(t,s)y(s)dtds - \frac{1}{2}\int_0^{\ell T}\int_0^{\ell T} y(t)\rho_n^{-1}(t,s)y(s)dtds. \quad (4.3)$$

Define $y_m(t)$ to be the output of the channel when the transmitted signal is $s_m(t)$, i.e.,

$$y_m(t) = \int_{\max\{t-LT,0\}}^{\min\{t,T\}} s_m(\tau)h_c(t-\tau)d\tau, \quad (4.4)$$

where it is assumed that $h_c(t) = 0$ for all $t \geq LT$ and for all $t < 0$. Note that $y_m(t) = 0$ whenever $t \notin [0,(L+1)T]$, and that by superposition,

$$y(t) = \sum_{k=0}^{\infty} y_{m(k)}(t-kT). \quad (4.5)$$

Defining

$$\beta_m(k) = \begin{cases} 1, \text{ if } m(k) = m \\ 0, \text{ otherwise,} \end{cases} \quad (4.6)$$

for $0 \leq t \leq \ell T$,

$$y(t) = \sum_{k=0}^{\ell}\sum_{m=1}^{M} \beta_m(k)y_m(t-kT). \quad (4.7)$$

Define

$$Z_{m,k,\ell}(t) = \int_0^{\ell T} \rho_n^{-1}(t,s) y_m(s-kT)ds, \tag{4.8}$$

$$\alpha_{m,k,\ell} = \int_0^{\ell T} r(t) Z_{m,k,\ell}(t)dt, \tag{4.9}$$

and

$$\gamma_{m_1,k_1,m,k,\ell} = \int_0^{\ell T} y_{m_1}(t-k_1T) Z_{m,k,\ell}(t)dt. \tag{4.10}$$

Note that from (4.8)-(4.10), $\gamma_{m_1,k_1,m,k,\ell} = \gamma_{m,k,m_1,k_1,\ell}$.

From the above definitions, (4.3) can be expressed as

$$\ell^*(r,\ell T) = \sum_{k=0}^{\ell} \sum_{m=1}^{M} \beta_m(k)\alpha_{m,k,\ell} - \frac{1}{2} \sum_{k=0}^{\ell} \sum_{m=1}^{M} \sum_{k_1=0}^{\ell} \sum_{m_1=1}^{M} \beta_m(k)\beta_{m_1}(k_1)\gamma_{m_1,k_1,m,k,\ell}. \tag{4.11}$$

Note that the Z's and the γ's are constants which may be computed *a priori* and that $\{\alpha_{m,k,\ell}: 1\leq m\leq M, 0\leq k\leq \ell\}$ is a set of sufficient statistics for $\{\hat{i}_k\}_{k=1}^{\ell}$. Furthermore, note that maximizing ℓ^* given by (4.11) with respect to $\{\beta_m(k): 1\leq m\leq M, 0\leq k\leq \ell\}$ is equivalent to maximizing ℓ^* given by (4.3) with respect to $\{i_k\}_{k=0}^{\ell}$

To be of any practical consequence for most digital communications applications, a recursive procedure for maximizing ℓ^* is necessary. In general, a recursive relationship (in ℓ) for ℓ^* is impossible unless either a recursive relationship exists for $\alpha_{m,k,\ell}$ and $\gamma_{m_1,k_1,m,k,\ell}$, or

$$\alpha_{m,k,\ell} = \alpha_{m,k} \tag{4.12}$$

and

$$\gamma_{m_1,k_1,m,k,\ell} = \gamma_{m_1,k_1,m,k}, \tag{4.13}$$

where the absence of the subscript ℓ is to be interpreted as being

independent of ℓ. The latter approach ((4.12) and (4.13)) is taken here, in analogy with the assumptions made by Ungerboeck [10]. With these assumptions, and using the symmetry of the γ's,

$$\ell^*(r,(\ell+1)T) = \ell^*(r,\ell T) + \sum_{m=1}^{M} \beta_m(\ell+1)\alpha_{m,\ell+1}$$
$$- \frac{1}{2}\sum_{m=1}^{M}\sum_{m_1=1}^{M} \beta_{m_1}(\ell+1)\beta_m(\ell+1)\gamma_{m_1,\ell+1,m,\ell+1}$$
$$- \sum_{k=0}^{\ell}\sum_{m=1}^{M}\sum_{m_1=1}^{M} \beta_m(k)\beta_{m_1}(\ell+1)\gamma_{m_1,\ell+1,m,k} \,. \tag{4.14}$$

Note that the number of computations needed to compute $\ell^*(r,(\ell+1)T)$ grows linearly with ℓ. Assuming that $\rho_n^{-1}(t,s)=0$ whenever $|t-s|>L_1T$, from (4.8)-(4.10) it can be shown that

$$Z_{m,k}(t) = 0 \text{ whenever } t\notin[(k-L_1)T,(L+L_1+k+1)T], \tag{4.15}$$

and

$$\gamma_{m_1,k_1,m,k} = 0 \text{ whenever } |k-k_1|>L+L_1. \tag{4.16}$$

Substituting (4.15) and (4.16) into (4.14), ℓ^* can be written as

$$\ell^*(r,(\ell+1)T) = \ell^*(r,\ell T)$$
$$+ \sum_{m=1}^{M} \beta_m(\ell+1)\{\alpha_{m,\ell+1} - \frac{1}{2}\gamma_{m,\ell+1,m,\ell+1}$$
$$- \sum_{k=\ell-L-L_1+1}^{\ell}\sum_{m_1=1}^{M} \beta_{m_1}(k)\gamma_{m,\ell+1,m_1,k}\}\,, \tag{4.17}$$

where use has been made of the properties of $\beta_m(k)$. The result (4.17) represents an extension of equation (27) of Ungerboeck [10] to general nonlinear modulation schemes.

Following Ungerboeck [10], a modified Viterbi algorithm can be applied to (4.17) to obtain a maximum-likelihood sequence estimate,

$\{\hat{i}_k\}_{k=1}^{\ell}$ for $\{i_k\}_{k=1}^{\ell}$. Recall that estimating $\beta_m(k)$ is equivalent to estimating $m(k)$ or i_k. Assume that $\{i_k\}$ is a coded sequence with μ_j the state of the coder after i_j has been transmitted. Assume further that the sequence of states $\{\mu_j\}$ is a Markov sequence, i.e., that $Pr(\mu_j | \mu_{j-1}, \mu_{j-2}, \ldots) = Pr(\mu_j | \mu_{j-1})$. Given μ_j and an allowable sequence $\{i_{j+1}, i_{j+2}, \ldots, i_{j+k}\}$, the state μ_{j+k} is uniquely determined. Consider the so-called survivor metric

$$\ell_{\ell,k}(\sigma_{\ell,k}) = \max_{\{\ldots, i_{\ell-k-1}, i_{\ell-k}\} \to \mu_{\ell-k}} \{\ell^*(r, \ell T)\}, \tag{4.18}$$

where $\sigma_{\ell,k} = (\mu_{\ell-k}; i_{\ell-k+1}, \ldots, i_\ell)$ is called the survivor state and the maximum is taken over all allowable sequences $\{i_j\}_{j=1}^{\ell-k}$ that put the coder into state $\mu_{\ell-k}$. Note that $Pr(\sigma_{\ell,k} | \sigma_{\ell-1,k}, \sigma_{\ell-2,k}, \ldots) = Pr(\sigma_{\ell,k} | \sigma_{\ell-1,k})$, i.e., that $\{\sigma_{\ell,k}\}$ is a Markov sequence. Associated with each $\sigma_{\ell,k}$ is at least one path history $\{\ldots, i^*_{\ell-k-1}, i^*_{\ell-k}\}$ which yields the maximum in (4.18). Ungerboeck [10] claims that this maximizing path history is unique. Define

$$J_{\ell,\ell+1} = -\sum_{m=1}^{M} \beta_m(\ell+1)\left(\frac{1}{2}\gamma_{m,\ell+1,m,\ell+1} + \sum_{j=\ell-L-L_1+1}^{\ell} \sum_{m_1=1}^{M} \beta_{m_1}(j)\gamma_{m,\ell+1,m_1,j}\right), \tag{4.19}$$

then

$$\ell^*(r, (\ell+1)T) = \ell^*(r, \ell T) + \sum_{m=1}^{M} \beta_m(\ell+1)\alpha_{m,\ell+1} + J_{\ell,\ell+1}. \tag{4.20}$$

Note that the only elements of $\{\beta_m(j)\}$ which appear explicitly in (4.19) and (4.20) are $\{\beta_m(j): 1 \le m \le M, \ell-L-L_1+1 \le j \le \ell+1\}$. Recall that a one-to-one relationship exists between i_j and $\beta_m(j)$. Consequently, letting $k = L+L_1+1$ and substituting (4.20) into (4.18),

$$\ell_{\ell+1,L+L_1+1}(\sigma_{\ell+1,L+L_1+1}) = \max_{\{\dots,i_{\ell-L-L_1-1},\,i_{\ell-L-L_1}\}\to\mu_{\ell-L-L_1}} \{\ell^*(r,\ell T) + J_{\ell,\ell+1}\} + \sum_{m=1}^{M} \beta_m(\ell+1)\,{}_{m,\ell+1}$$

$$= \sum_{m=1}^{M} \beta_m(\ell+1)\alpha_{m,\ell+1} + \max_{\sigma_{\ell,L+L_1+1}\to\sigma_{\ell+1,L+L_1+1}}\{\ell^*(r,\ell T) + J_{\ell,\ell+1}\}$$

$$= \sum_{m=1}^{M} \beta_m(\ell+1)\alpha_{m,\ell+1} + \max_{\sigma_{\ell,L+L_1+1}\to\sigma_{\ell+1,L+L_1+1}}\{\ell_{\ell,L+L_1+1}(\sigma_{\ell,L+L_1+1}) + J_{\ell,\ell+1}\}. \tag{4.21}$$

For each ℓ, survivor metrics and path histories must be calculated for all possible states $\sigma_{\ell,L+L_1+1}$. One would expect that the further one looks back from ℓ, all path histories would tend to be identical. In practice, one would choose the path history with the largest survivor metric for some $\ell-L^*$. If one chooses L^* large enough, then the difference in performance between allowing an unbounded delay and a delay of L^* is negligible. Apparently, analytical guidelines for the choice of L^* is an open issue.

B. Adaptive Maximum-Likelihood Sequence Estimation

In order to utilize the maximum-likelihood sequence estimation algorithm represented by (4.21), the channel, $\{h_c(\tau): 0\leq\tau<LT\}$, and the inverse kernel $\{\rho_n^{-1}(t,s):|t-s|\leq L_1T\}$ are needed. In this section, adaptive techniques for approximating h_c and ρ_n^{-1} are presented.

First, consider adaptively estimating the channel, h_c. Recall that $r(t) = y(t) + n(t)$, and that

$$y(t) = \sum_{k=0}^{\infty} y_{m(k)}(t-kT) = \sum_{k=0}^{\infty} \sum_{m=1}^{M} \beta_m(k)y_m(t-kT). \tag{4.22}$$

If the channel were known and if the transmitter and receiver clocks were synchronized, then the signal components $\{y_m(\cdot):1\leq m\leq M\}$ could be generated at the receiver using (4.4). In this case, y(t) could be estimated as

$$y(t) = \sum_{k=0}^{\infty} \sum_{m=1}^{M} \beta_m(k)y_m(t-kT), \tag{4.23}$$

where $\{\hat{\beta}_m(k)\}$ is an estimate of $\{\beta_m(k)\}$ obtained from the modified Viterbi algorithm.

Since the channel is unknown, consider an NL-tap transversal filter with tap spacing D, tap weights $\{w_i\}_{i=1}^{NL}$, and input $s_m((j-n_c)D)$, where n_c is an integer used to denote the unknown timing relationship between the transmitter and receiver clocks. Let ND = T and consider

$$\hat{y}_m(jD) = \sum_{i=1}^{NL} s_m((j-i+n_c+1)D)w_i \tag{4.24}$$

as an estimate of $y_m(jD)$. With the adaptive direct channel model of Section III-B as motivation, consider

$$\hat{y}(jD) = \sum_{k=0}^{\infty} \sum_{m=1}^{M} \hat{\beta}_m(k)\hat{y}_m(jD-kT), \tag{4.25}$$

as an estimate of $y(jD)$, with $\{w_i\}_{i=1}^{NL}$ chosen to minimize

$$\xi(W) = N^{-1} \sum_{j=1}^{N} E((r(jD)-\hat{y}(jD))^2), \tag{4.26}$$

where

$$W = (w_1, w_2, \ldots, w_{NL})'. \tag{4.27}$$

Define

$$S_{m,j,k} = (s_m((j+n_c)D-kT), \ldots, s_1((j-NL+n_c+1)D-kT))' \tag{4.28}$$

and

$$Z_j = \sum_{k=0}^{\infty} \sum_{m=1}^{M} \hat{\beta}_m(k)\, S_{m,j,k}. \tag{4.29}$$

Then

$$\hat{y}(jD) = W'Z_j, \tag{4.30}$$

and the function to be minimized is

$$\xi(W) = \frac{1}{N} \sum_{j=1}^{N} E(r_j^2) - \frac{2}{N} W' \sum_{j=1}^{N} E(r_j Z_j) + \frac{1}{N} W' \sum_{j=1}^{N} E(Z_j Z_j')W, \tag{4.31}$$

where $r_j = r(jD)$.

Assuming that $\{i_k\}$ and $\{n(t)\}$ are jointly wide-sense stationary, $E(r^2(jD))$ is periodic (in j) with period N. If $\hat{\beta}_m(k) = \beta_m(k)$, then $E(r_j Z_j)$, $E(\hat{y}^2(jD))$, and $E(Z_j Z_j')$ are also periodic (in j) with period N. It is assumed in the following that these periodicities are present.

If

$$R = N^{-1} \sum_{j=1}^{N} E(Z_j Z_j') \tag{4.32}$$

is positive definite and

$$P = N^{-1} \sum_{j=1}^{N} E(r_j Z_j) \tag{4.33}$$

is nonzero, then the unique minimizing weight vector, w_o, is given by $w_o = R^{-1}P$.

Let $\{h_i\}_{i=1}^{NL}$ be the weighting sequence for a transversal filter model of the channel, h_c, with tap spacing D. Then $y_m(jD)$ (from (4.4)) can be approximated as

$$y_m(jD) \approx \sum_{i=1}^{NL} s_m((j-i+1)D)h_i. \tag{4.34}$$

Defining

$$H = (h_1,h_2,\ldots,h_{NL})', \tag{4.35}$$

then

$$y_m(jD - kT) \approx S'_{m,j-n_c,k}H, \tag{4.36}$$

and

$$y(jD) \approx \sum_{k=0}^{\infty} \sum_{m=1}^{M} \beta_m(k)S'_{m,j-n_c,k}H. \tag{4.37}$$

Now, assuming that $\beta_m(k) = \hat{\beta}_m(k)$ and that $\{\beta_m(k)\}$ is approximately independent of $\{n(t)\}$,

$$y(jD) \approx Z'_{j-n_c}H, \tag{4.38}$$

and

$$E(r_jZ_j) \approx E(Z_jZ'_{j-n_c})H. \tag{4.39}$$

Consequently, from (4.32), (4.33), and $w_o = R^{-1}P$, w_o is an approximate solution to

$$\sum_{j=1}^{N} E(Z_jZ'_j)W = \sum_{j=1}^{N} E(Z_jZ'_{j-n_c})H. \tag{4.40}$$

Define $\{z_\ell\}$ by $z_\ell = (Z_\ell)_1$. By assumption, $\{z_\ell\}$ is a "wide-sense cyclostationary process," i.e., $E(z_\ell z_{\ell+u})$ is periodic (in ℓ) with period N. Defining

$$\rho_z(u) = \sum_{\ell=1}^{N} E(z_\ell z_{\ell+u}), \tag{4.41}$$

and letting $w_j = (W)_j$, (4.40) may be expressed as

$$\sum_{j=1}^{NL} \rho_z(i-j)w_j = \sum_{j=1}^{NL} \rho_z(i-j-n_c)h_j, \quad i=1,2,\ldots,NL. \tag{4.42}$$

Equation (4.42) suggests that $\{w_j\}$ is a shifted version of $\{h_j\}$. For example, suppose that $NL >> n_c > 1$, and that $h_j = 0$ for $NL-n_c+1 \leq j \leq NL$. By assumption, the solution $\{w_j\}$ to (4.42) is unique. Consequently, since

$$w_j = \begin{cases} h_{j-n_c} & , \; n_c+1 \leq j \leq NL \\ 0 & , \; 1 \leq j \leq n_c \end{cases} \tag{4.43}$$

satisfies (4.42), it must be the unique solution. These remarks offer strong justification to the claim that if $|n_c| << NL$, then $w_o = R^{-1}P$ is approximately a shifted version of the discrete-time channel model, H. The resulting shift is, of course, desired in order to line up the receiver-generated estimate of the received process with the actual received process. This property is quite similar to that of the direct channel model of Section II-B, except the present scheme exhibits no intersymbol interference.

Suitable algorithms for estimating $w_o = R^{-1}P$, with R and P given by (4.32) and (4.33), are represented by

$$W_{n+1} = W_n + \mu_n(P_n - F_nW_n), \tag{4.44}$$

where

$$F_n = K_n^{-1} \sum_{\ell=n-K_n+1}^{n} Z_\ell Z_\ell', \tag{4.45}$$

and

$$P_n = K_n^{-1} \sum_{\ell=n-K_n+1}^{n} r_\ell Z_\ell. \tag{4.46}$$

Such algorithms fit the general framework presented in Chapter II,

and hence, the convergence results of Chapter II are applicable to (4.44)-(4.46). If $\hat{\beta}_m(k) \equiv \beta_m(k)$, then Theorem 2 of Chapter II is applicable to obtain rather mild conditions for which $W_n \overset{a.s.}{\rightarrow} w_o$. The occurrence of $\hat{\beta}_m(k)$ in (4.29) represents what is commonly called "decision-feedback." Such schemes always have a finite probability of a "runaway."

Now consider the problem of estimating the inverse kernel, ρ_n^{-1}. Suppose that a time-invariant approximate whitening filter, $h_w(\tau)$, for $\{n(t)\}$ exists. Then, denoting the output of the whitening filter by $n^*(t)$, the spectral density of $n^*(t)$ by $S^*(f)$, the spectral density of $n(t)$ by $S_n(f)$, and the transfer function for $h_w(\tau)$ by $H_w(f)$, we have

$$1 \approx S^*(f) = S_n(f) \, |H_w(f)|^2. \tag{4.47}$$

Consequently,

$$\rho_n^{-1}(\cdot,\cdot) \approx F^{-1} \{\frac{1}{S_n(f)}\} = F^{-1}\{|H_w(f)|^2\}, \tag{4.48}$$

where F^{-1} denotes inverse Fourier transform. From (4.48) it is apparent that h_w convolved with itself is an approximation to ρ_n^{-1}. Hence, via (4.48), an approximation of h_w will result in an approximation to ρ_n^{-1} as well.

Recall that $r(t) = y(t) + n(t)$. Using $\hat{y}(jD)$ from (4.25), an estimate of $n(jD)$ is given by

$$\hat{n}(jD) = r(jD) - \hat{y}(jD). \tag{4.49}$$

Techniques analagous to those proposed, e.g., by Widrow *et. al.* [13] will now be used to obtain an adaptive transversal filter approximation to h_w. Define

$$\hat{N}_\ell = (\hat{n}((\ell-1)D), \hat{n}((\ell-2)D),\ldots,\hat{n}((\ell-p)D))' , \tag{4.50}$$

and

$$W = (w_1, w_2, \ldots, w_p)'. \tag{4.51}$$

Define

$$n_\ell^* = \hat{n}(\ell D) - W'\hat{N}_\ell, \tag{4.52}$$

and consider choosing W to minimize

$$\xi(W) = E((n_\ell^*)^2). \tag{4.53}$$

Defining

$$R_{nn} = E(\hat{N}_\ell \hat{N}_\ell'), \tag{4.54}$$

and

$$P = E(n_\ell^* \hat{N}_\ell), \tag{4.55}$$

the minimizing weight vector, w_o, is given by

$$w_o = R_{nn}^{-1} P, \tag{4.56}$$

assuming that R_{nn} is positive definite. The main idea is that $w_o'\hat{N}_\ell$ is the best (in the minimum mean-squared error sense) linear predictor of $\hat{n}(\ell D)$ based on the previous p samples, $\hat{\mathbf{N}}_\ell$. Subtracting this prediction of $\hat{n}(\ell D)$ from $\hat{n}(\ell D)$, should then tend to "whiten" the residual, n_ℓ^*. The resulting approximation to K_n^{-1} is a transversal filter with unit pulse response cg_u, with

$$g_u = \begin{cases} 1, & u=0 \\ -2w_1, & u=1 \\ -2w_u + \sum_{i=1}^{u-1} w_i w_{u-i}, & 2 \le u \le p \\ \sum_{i=u-p}^{p} w_i w_{u-i}, & p+1 \le u \le 2p \\ 0, & \text{elsewhere,} \end{cases} \tag{4.57}$$

where c is a constant used to normalize (4.52) so that n_ℓ^* has unit variance. Of course, a suitable estimate for c is

$$\hat{c}_\ell = \Big(\frac{1}{K_\ell}\sum_{j=\ell-K_\ell+1}^{\ell}(n_j^*)^2\Big)^{-1} , \tag{4.58}$$

where $K_\ell=1$, K, or ℓ, for example.

Suitable algorithms for approximating w_o given by (4.56) are represented by

$$W_{n+1} = W_n + \mu_n(P_n - F_n W_n), \tag{4.59}$$

with

$$F_n = K_n^{-1}\sum_{j=n-K_n+1}^{n} \hat{N}_\ell \hat{N}_\ell' \tag{4.60}$$

and

$$P_n = K_n^{-1}\sum_{j=n-K_n+1}^{n} n_\ell^* \hat{N}_\ell \; . \tag{4.61}$$

In case $\hat{n}(\ell D) \equiv n(\ell D)$, then from Chapter II, Theorem 7, $W_n \overset{a.s.}{\to} w_o$ provided that for some $\nu > \frac{1}{4}$, $u^\nu |\rho_n(t,t+u)|$ is uniformly bounded for all real t and all non-negative integers u, and that $\{\mu_n\}$ satisfies Condition B3.

V. CONCLUSION

Several adaptive communication systems have been presented and analyzed. The analysis has included convergence properties and various kinds of "self-synchronization." In many applications, e.g., transmission of digital data over the undersea channel, intersymbol interference becomes an important problem. For such applications, the adaptive maximum-likelihood sequence estimation schemes presented in Chapter IV are especially appropriate.

The results presented in Chapter IV, which are believed to be new, require additional analytical investigations as well as simulation studies, before corresponding practical systems can realistically be proposed. Of primary importance are the convergence properties of decision-directed algorithms and suitable initialization procedures. Decision-directed algorithms suffer from a phenomenon called a "runaway." A runaway occurs when the detector makes a sequence of erroneous decisions, which degrades the parameter estimates, which in turn further degrade the detector performance. The only analytical results treating this problem of which this author is aware are those of Davisson and Schwartz [16]. A possible method of reducing the probability of a runaway is to incorporate a "null-zone" detection scheme into the decision-feedback loop. A combination of the null-zone decision-feedback scheme of Gitlin and Ho [17] and the results of Chapter IV could well lead to a practical "high performance" receiver for "nonlinear" modulation schemes.

Closely related to the techniques presented in Chapter III are

adaptive reference estimators [19] used in conjunction with a pre-distorted-replica correlation receiver [18]. Such techniques, which make use of decision-feedback strategies, suffer a degradation in performance due to intersymbol interference in a manner quite similar to the schemes presented in Chapter III. Furthermore, such schemes exhibit a nonzero probability of a runaway, as do other decision-feedback strategies.

REFERENCES

[1] D. C. Farden, "Stochastic Approximation with Correlated Data," Technical Report, Contract Number N66001-75-C-0224, April 1976, also submitted to Annals of Statistics, April 1976.

[2] T. P. Daniell, "Adaptive Estimation with Mutually Correlated Training Sequences," IEEE Trans. Syst. Sci. Cybernet., SSC-6, Jan. 1970, pp. 12-19.

[3] J. K. Kim and L. D. Davisson, "Adaptive Linear Estimation for Stationary M-dependent Processes," IEEE Trans. Inform. Theory, IT-21, Jan. 1975, pp. 23-31.

[4] D. J. Sakrison, "Stochastic Approximation: A Recursive Method for Solving Regression Problems," Advances in Comm. Sys. 2, A. V. Balakrishnan, editor, Academic Press, New York and London, 1966, pp. 51-106.

[5] D. C. Farden, "Stochastic Approximation with Correlated Data," Ph.D. dissertation, Dept. of Electrical Engineering, Colorado State University, Fort Collins, also ONR Technical Report No. 11, May 1975.

[6] R. J. Serfling, "Moment Inequalities for the Maximum Cumulative Sum," Ann. Math. Statist., 41, 1970, pp. 1227-1234.

[7] R. J. Serfling, "Convergence Properties of S_n Under Moment Restrictions," Ann. Math. Statist., 41, 1970, pp. 1235-1248.

[8] P. Butler and A. Cantoni, "Noniterative Automatic Equalization," IEEE Trans. Commun., COM-23, June 1975, pp. 621-633.

[9] G. D. Forney, Jr., "Maximum-Likelihood Sequence Estimation of Digital Sequences In the Presence of Intersymbol Interference," IEEE Trans. Inform. Theory, IT-18, May 1972, pp. 363-378.

[10] G. Ungerboeck, "Adaptive Maximum-Likelihood Receiver for Carrier-Modulated Data-Transmission Systems," IEEE Trans. Commun., COM-22, May 1974, pp. 624-636.

[11] R. W. Lucky, "A Survey of the Communication Theory Literature: 1968-1973," IEEE Trans. Inform. Theory, IT-19, Nov. 1973, pp. 725-739.

[12] B. Widrow, "Adaptive Filters," in Aspects of Network and System Theory, edited by R. E. Kalman and N. DeClaris, Holt, Rinehart and Winston, New York, 1970, pp. 563-587.

[13] B. Widrow, *et. al.*, "Adaptive Noise Cancelling: Principles and Applications," Proc. IEEE, vol. 63, Dec. 1975, pp. 1692-1716.

[14] F. R. Magee and J. G. Proakis, "Adaptive Maximum-Likelihood Sequence Estimation for Digital Signaling in the Presence of Intersymbol Interference," IEEE Trans. Inform. Theory, vol. IT-19, pp. 120-124, Jan. 1973.

[15] S.U.H. Qureshi and E. E. Newhall, "An Adaptive Receiver for Data Transmission over Time-Dispersive Channels," IEEE Trans. Inform. Theory, vol. IT-19, pp. 448-457, July 1973.

[16] L. D. Davisson and S. C. Schwartz, "Analysis of a Decision-Directed Receiver with Unknown Priors," IEEE Trans. Inform. Theory, vol. IT-16, pp. 270-276, May 1970.

[17] R. D. Gitlin and E. Y. Ho, "A Null-Zone Decision Feedback Equalizer Incorporating Maximum Likelihood Bit Detection," IEEE Trans. Commun., COM-23, pp. 1243-1250, Nov. 1975.

[18] D. C. Farden, F. N. Cornett, and L. L. Scharf, "Communication Receiver Studies," Final Report for Contract No. N66001-74-C-0035, Naval Undersea Center, San Diego, California.

[19] D. C. Farden and L. L. Scharf, "Communication Receiver Studies - II," Research Proposal, March 1974.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

(6) Adaptive Baseband Digital Communication and Applicable Convergence Results.

5. TYPE OF REPORT & PERIOD COVERED

(9) Final Report.

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

(10) David C. Farden

8. CONTRACT OR GRANT NUMBER(s)

(15) N66001-75-C-0224, ARPA Order-1872

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Department of Electrical Engineering
Colorado State University
Fort Collins, Colorado 80523

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

(12) 53 p.

11. CONTROLLING OFFICE NAME AND ADDRESS

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

12. REPORT DATE

(11) May 1976

13. NUMBER OF PAGES

50

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

Naval Undersea Center
San Diego, California 92132

15. SECURITY CLASS. (of this report)

Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

adaptive signal processing, adaptive communication systems, LMS algorithm, stochastic approximation, almost sure convergence, moment conditions, covariance decay-rate conditions, correlated data, cyclostationary processes, self-synchronization, maximum-likelihood sequence estimation, intersymbol interference

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Several baseband adaptive communication systems are presented and analyzed with regard to "self-synchronization" properties, intersymbol interference, and convergence properties. Recent convergence results, the proofs of which are contained in a companion report, are applied to provide extremely mild "covariance decay-rate conditions" for which the algorithms treated converge with probability 1. Of special interest are the convergence results treating correlated cyclostationary training data. Recent results on maximum- OVER

406 434

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Cont'd

likelihood sequence estimation are extended to treat the detection of general "nonlinearly modulated" digital data over linear dispersive channels and nonwhite additive noise. Adaptive techniques for training the new detector structure are proposed for use when the channel and/or the noise covariance function are unknown.